ISSN:0975-9646

Dipak R. Pardhi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5512-5517

Personalized Image Search Using Simple Case of One Word-Based Queries-An Overview

Prof. Dipak R. Pardhi[1], Mr. Lalitkumar B. Borase[2]

Department of Computer Engineering, Godavari College of Engineering, NMU, Jalgaon (MH)
GF's Godavari College of Engineering, Jalgaon (MH), India

***Abstract*— In today's web search generates a large amount of data and search contents. Text based searching is most popular technique to search content based on available information. In social websites like Flicker, YouTube, Face book they provide features to users that create information, share information and allows the users to search information based on created tag, comment and annotation. This Create the user-generated metadata in very large amount which can be utilized for management and information or media retrieval.**

It's found that use of tag, comment, and annotated data is convoluted in personalized search system. With Personalize search technique is improved by using user preferences and query-related search intention into user-specific topic spaces. In this paper the given model is tested for single word based query and showing satisfactory result.

***Keywords*— Personalize search, Non-Personalize search, User preferences, Image annotation.**

I. INTRODUCTION

This Social networking website has been hosting huge number of images from last few years the growth of social web images data is huge amount. User's text based searching and generating content is basically using single keyword that depends on interest of users and a large quantity of metadata in the form of tags and ratings, which can be used to improve web search and personalization. Personalized Image Search is the basic need of every web surfer/user seeking for images of his/her interest. In Personalization searching technique uses previous history of user as relevance feedback to improve the precision of system. Personalization searching technique usually uses databases to store the search history, previously visited web pages, email id of the user to exploit the search intent. Major challenge face by these personalized search system are –

1) Large amount of queries to search engines are short and ambiguous, and other users may have completely dissimilar information needs and goals under the same query. 2) There exist different forms of metadata, such as descriptions, annotation and ratings. So how to model other metadata for an overall system so it's also another challenge.

To address the problem of generating the personalized search and recommended the tags, annotation and description for a given query by the user, the proposed system provide solution using ternary relation among user, image and tag technique, and customize Ranking base Multi-correlation tensor factorization (RMTF) service.

A given user query a personalized image search system tries to recommended the images that are closely relevant to this search intent. The personalized search approach in this paper considers the users preference for ranking the search results. The basic assumption is that users tagging actions reflects their personal relevance judgment. But the fact is that user-specified annotations are not sufficient for building their Preference Profile. So the task of annotation prediction is achieved using Tensor Factorization using HOSVD method. There are three entities in photo sharing websites User, Image and Tag respectively, which forms ternary relation between them. The users tagging patterns are-

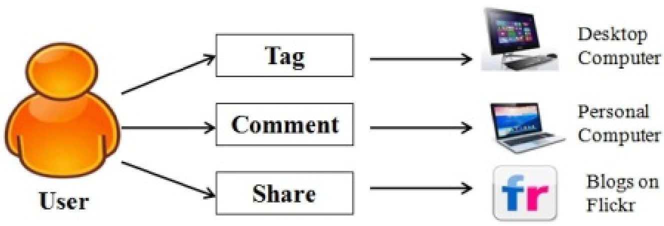

Figure1. Tagging Data Pattern

The number of tags is produced by Tensor Factorization using HOSVD method is considered as tag Vocabulary for that user. One query may correspond to more than one tags in the tag vocabulary that means this may lead to one-to-many relation between user query and tags. Hence the mapping between the query and tags is achieved using User-Specific Topic modeling, where user query is mapped to one topic from set of user-specific topics. If user u1 has a principle interest on topic j and term X has a high probability in topic j, when user searches X, the query will have a high proportion on users topic j.

In personalized image search results from the search engines. Personalized search results consider as both user query relevance and user preference, so the personalized outcome from a laptop lover rank the laptop images on the top, as per shown in figure 2. This provides a basically two step solution scheme. Most of the existing work [2]–[5] follow this scheme and decompose personalized search into two steps: computing the non-personalized consequence score between the query, the document, and computing their personalized score by estimate the user's preference over the document. After this, a merge operation is conducted to generate a ranked list. It suffers from two problems. 1) The interpretation is less straight and it's not that much persuaded. The intuition of personalized search is to rank the returned documents by estimating the user's preference over documents in the particular queries. Over that they can directly analyzing the user query document relationship, the existing system scheme approximates it by separately computing a user query document relevance score and a user document significance score. 2) How to determine the merge strategy is not trivial. It research, simultaneously considers the normally a weighting parameter will be optimized to balance the two scores, or the learnt user preference is used to re-rank the query significance-based original list.

Figure2. Personalized Image search

To investigate on user preference and perform user modeling, the most popular social activity of tagging is considered. Collaborative tagging has popular for sharing and organizing resources, leading to a large amount of user-generated annotations. Photo sharing websites, such as Flickr, YouTube, and Face book are allowing users as owners, taggers, or commenter's for their provided contents to interact and collaborate with each other in a social media. Various researchers have investigated the applicability of social annotations to improve web search [6]–[10]. Recently, social annotations are employed for automatic evaluation of personalized search [2], [11], [3].

A primary assumption is that, the users' tagging actions reflect their personal relevance judgment. For example, if a user assign tag "Laptop" to an image, it is probable that the user will consider this image as relevant if user issues "actor" as a query. The intention in that if the user's annotations to the images are available, it can directly estimate the user's preference under certain queries, and the fact is that the original annotations available are not enough for user preference mining. Therefore, it transfers the problem of personalized image search to user's annotation forecast. Moreover, as queries and tags do not follow simple one-to-one relationship, it builds user-specific topic spaces to develop the relations between queries and tags.

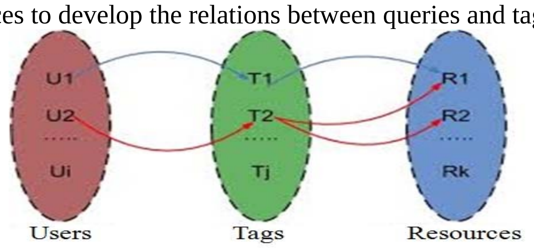

Figure3. Social Tagging Concepts

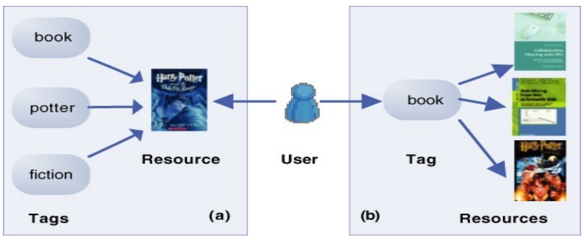

Figure4. Example of tag, blog, comment and share

A basic assumption is that, the users is tagged actions is to reflect their personal relevance understanding. For example, if a user tagged "book" to an image, it is probable that the user will believe this image as relevant if he/she issues "book" as a user query. Illustrated by this, the intuition of this research is that if the user's annotations to the images are available, they can estimate directly to the user's preference under certain queries.

II. Literature survey

In recent years most of the work has been done on improving the web search. In that firstly described the survey of some foregoing efforts on Personalize Search, and find out the boundaries of these works in terms of the user profiling and user interest that is significance measurement and improving results.

P. Symeonidis, A. Nanopoulos, Y. Manolopoulos stated that Automated Analysis of Interests and Activities[12], Here the information about the searcher is used to infer an implicit goal or intent. The search related information such as previously issued queries and previously visited Web pages, and on other information such as credentials and email the user has read and created can be explored. The completely constructed user profile as a form of significance feedback can achieve better performance than explicit relevance feedback and can improve on Web search. With this approach to personalization, there is no need for the user to specify or maintain a profile of interests.
Drawback: Need to use a wide range of implicit user activities over a long period of time to develop a contained user profile. This profile is used to re-rank Web search results employing a relevance feedback framework.

Dongyuan Lu, Qiudan Li stated that Personalized Search on Searchers Preference Prediction[5]: This personalized search model assists users in obtaining interested photos on Flickr, by exploiting the favourite marks of the searchers friends to predict the searchers preference on the returned photos. This model utilizes a co-clustering method to extract latent interest scope from user's implicit interests, and employs a discriminative learning method to predict searchers preference on the returned photos.
Drawback: Typically, users are interested in more than one field, and the searcher may share different interests with dissimilar friends. The variety of users contained interests can be mined and encoded into the latent interest scope. Friends may contribute another way to searchers preference prediction according to the submitted query and the interest distribution.

P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos stated that Personalized Search Based on Social Annotations [13]: This system explores the use of social annotations to improve web search. The web search optimized by using social annotations from the following two aspects: 1) Similarity ranking: The similarity between a query and a web page. For example, the top 5 annotations of Amazons homepage in Delicious are shopping, books, amazons, music and store, which show the page or even the whole website exactly. These annotations provide a new metadata for the similarity estimation between a query and a web page. 2) Static ranking: The amount of annotations assigned to a page indicates the popularity of web pages using social annotations.

Drawback: First, the user submitted queries may not match any social annotation. Second, many web pages may have no annotations. Annotation ambiguity is another problem concerned with Similarity Ranking, i.e., the similar terms to the query terms while fail to disambiguate terms that have more than one meanings. For example, ticket may refer to either airplane ticket or concert ticket, and terms with these two different meanings will be mixed up.

P. Symeonidis, A. Nanopoulos, Y. Manolopoulos stated that Recommendations Based on Ternary Semantic Analysis [12] Social tagging systems (STSs) can provide three different types of recommendations: They can suggest 1) tags to users, base on what tags other users have used for the same items, 2) items to users, based on tags they have in familiar with other similar users, and 3) users with common social interest, based on familiar tags on similar items. However, users may have dissimilar interests for an item, and items may have multiple facets. In this system data are modelled by a 3-order tensor, on which multi-way latent semantic analysis and dimensionality decrease is performed using both the Higher Order Singular Value Decomposition (HOSVD) method and the Kernel SVD smoothing technique.

Drawback: Need to apply different weighting methods for the initial construction of a tensor to improve the overall performance of web search

Based on Xiaoou Tang, Ke Liu, Jingyu Cui Capturing User Intention for One-Click Image Search [14]: It is difficult to interpret user's intention only by query words and this leads to ambiguous and noisy search results which are far from satisfactory. Hence to overcome this problem an approach is proposed which requires the user to click on one query image with least attempt and images from a pool retrieved by text based search are re-ranked based on both visual and textual content.

Drawback: User intention cannot be well expressed by a single query image. The user may be interested in only part of the image. In those cases, more user interactions, such as labelling the regions that the user thinks are more important have to be allowed. However more user burden has to be added.

III. RELATED WORK

In the recent years, wide efforts have been focus on personalized search. Regarding the explicit user profile, relevance response, user history data (browsing log, click-through data, tag and social annotations etc.) perspective information is time and place, etc. and social network are also broken for the implementation there are two basic stages

1. Query refinement and 2. Result processing

A. Ranking Based Multi-correlation Tensor Factorization

Its present the algorithm for annotation forecast. Table - 1 list the types of notations used in that research. There are three types of entities, they have to share image to many websites like facebook, picasa, flickr etc. The tagged information can be viewed as a set of triplets. Predicting the users comments to the images are related to reconstructing the user-tag-image in ternary interrelations. The low-rank approximation is performing to use Tucker decomposition [16] in a general tensor factorization model. The model RMTF is proposed to designed as objective function. It firstly introduces a ranking based optimization scheme for demonstration of the tagging data or tag the image to perform an improved dominance images. The users may among various inter relation images and tags are utilized as the efficiency constraints to take on it.

TABLE I
LIST OF KEY NOTATIONS

Symbol	Description
Y, C	User-image-tag tensor and core tensor
U, I, T	represent user, image tag factor matrices
U, I, T	sets of users, images and tags, respectively
u, i, t	represent user, image, tag index
u, i, t	represent user, image, tag feature vectors

B. Ranking Based optimization scheme

In this type of situation of social image tagging, the semantics of encoding every unseen data as 0 are incorrect, this is illustrated with example.

The scenario is that *user3* has not given any tag to *image2* and *image4* does not mean *user3* considering every tag is bad for set forth the images. Maybe he/she does not want the image to tag or it has no chance to see the fifth image and *user1* annotates *image1* with only the third *tag*. It is also indefensible to assume that further tags should not be comment on to the image, as many concepts may be absent in the user generated tags and individual user may not be well-known to all the related tags in the large tag dictionary. According to the optimization function *0/1 scheme* tries to forecast 0 for both cases. The above two issues, in this research, it present a ranking optimization scheme which naturally takes the user tagging behaviours into reflection. First of all they have to note that only the flexible dissimilarity is important and appropriate to the numerical values of 1 and 0 is unnecessary.

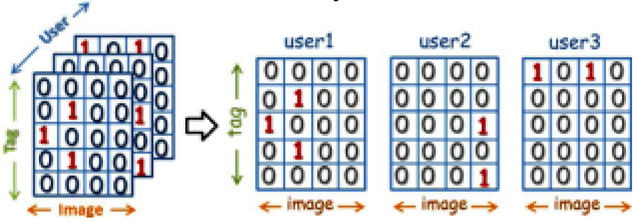

Figure 5(a). Tagging data interpretation using 0/1 scheme.

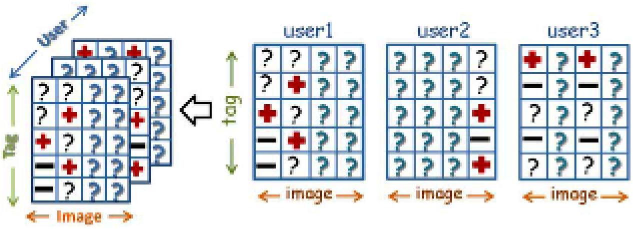

Figure 5(b). Tagging data analysis using Ranking scheme

C. Multi-correlation Smoothness Constraints

Photo or image sharing websites distinguish from other social tagging systems by its features of self-tagging. The most images are only tagged by their users or owners. In addition to the ternary interrelations, they collect multiple intra-relations among users, images and tags. These intra-relations represent the affinity graphs it's assume that two items with high affinities must be map close to each other in the learnt factor subspaces. In that they construct the tag affinity graph, and then integrate them into the tensor factorization framework. The user affinity graph and image affinity graph are constructed based on statistics of conjoined groups and visual similarity, respectively.

D. Tag affinity graph

The ranking based optimization scheme, construct the tag affinity graph based on the tag semantic relevance and context relevance tag and is simply encoded by their weighted co-occurrence in the image collection for tag semantic relevance, they estimate the semantic relevance between tags and based on their WordNet distance. WordNet is Lexical Database or it is a hierarchy and using the information content values of the concepts. This approach [17][18] takes both of the concept and their common ancestor in the calculation of similarity.

Jiang Conrath [17] measure gives semantic distance rather than similarity or relatedness.

$$Dist(t_1,t_2) = IC(t_1 + (t_2) - 2 * IC(t)$$

Where t is the concept providing the maximum information content shared by t_1 and t_2 in the taxonomy. This distance measure can be converted to a similarity measure by taking the multiplicative inverse of it:

$$Sim(t_1,t_2) = \frac{1}{Dist(t_1,t_2)}$$

Thus Sim (t_1, t_2) gives the similarity between concept t_1 and concept t_2.

E. User Specific Topic Modeling

TABLE II
EXAMPLE OF USER-SPECIFIC TOPICS

User A	Topic 1	military, aircraft, navy, iraq, artillery
	Topic 2	apple, computer, art, girl, cellphone
	Topic 3	Athelete, baseball, acto, sports, art, film
	. . .	. . .
User B	Topic 1	Buddha, budhist, temple, religion, asia
	Topic 2	Blossoms, blooms, nature, macro, flower
	Topic 3	Airplane, boeing, aircraft, airport, jet
	. . .	. . .

The rebuild user-tag-image ternary combination, so the personalized image search is to perform directly. When user submits a query, the rank of image is inversely relative to the likelihood of annotating with tag submitted query. The queries and tags do not trail one-to-one relationship query usually corresponds the tag terminology is to several related tag. In the case, the query-tag correspondence differs from user owner to user. Therefore, they build topic spaces for each user to use this user-specific one-to-many correlation. the analysis on a Flickr dataset of 270-K images that the usual number of annotated images per user is just 30. From the user-specific topics, it can see-

1) User's interest profile, for e.g., user is likely to be a computer device who also likes laptop, palmtop and desktop, while user is intense at religion and interests in personal computer and workstations.

2) The same tag may have different topic subsequent distributions for different users, e.g., for user, car occurs frequently in an automobile-related topic, while for user, car returns to its plain sense of vehicle.

IV. PERSONALIZED BASED RESULTS

TABLE III
TESTED SET STATISTICS

Testing Set	User	Query	Images tagged/ favorited	Tags annoted
NUS-IDE15_A10_30	30	11	253	14,148
NUS-WIDE15_A100	30	18	4,566	319,702
NUS-WIDE15-10_30	30	15	233	5,015
NUS-WIDE15_F100	19	15	3,214	19,254

In the research community of personalized search, is not an easy task since relevancy conclusion can only be evaluated by the users or the searchers themselves. The most accepted approach is user study, where participants are asked to judge the search results. Clearly this approach is so expensive. In addition, a main problem for user study is that the results are likely to be prejudiced as the participants know that they are being tested. Another broadly used approach is by user query logs history. However, these requirements a large scale real search logs, which is not available to most of the researchers. Social sharing websites present prosperous resources that can be broken for personalized search is calculate. User's social behaviours, such as rating as score the image, tagging as tag the image or document to other users and commenting the image, indicate the user's interest and user preference in a specific document.

Figure 7(a). Non-personalized search results using Topic-based method.

Figure 7(b). Personalized search results using Topic-based method.

Figure 8(a). Non-personalized search results using RMTF LDA.

Figure 8(b). Personalized search results using RMTF LDA.

The non-personalized results and the personalized results of User A and User B, It can also considering the query relevance and user information, to the proposed system (RMTF – Ranking based multi-correlation tensor factorization) and (LDA- Latent Dirichlet Allocation) captures the user's preference under certain topics. As a result of mapping art, nature, flowers, etc to Topic-2 of Table II, the top search results for user A mainly focus on blossom, blooms etc. While, for user B, the above search results are basically military related, which coincides with user B preference. For the baseline method which is having the separate query relevance and user preference, sometimes it's very hard to interpret the search results. For example in figure 7(a), figure. 8(a) is non-personalized outcome and fig. 7(b), fig. 8(b) personalized outcome with query "computer" having common understanding to the variant users and methods (topic based and RMTF-LDA) incorporating user information will make confusing search results. There are literatures are discussing the issue about when to perform personalization. Benefit of personalization is highly dependent on the ambiguity of the user query. Since there is no conclusion to this problem, in this research they focus on the problem of how to perform personalization and debate of when to perform personalization is afar the scope of this research.

V. Conclusion and Feature Work

The internet is huge source for receiving useful information. Search engines try to provide better solution for user's difficulty, by allowing them to indicate a query and providing the images that assure them. It is most complex for the user to select the image among the outcome shown by search engine.

In that system to use the user's social behaviour for personalized image search using single word based queries. These activities include annotations, tag, description and the participation of user in groups of interest. The query relevance and user preference are mutually at a time combined into the final rank list in order to achieve result as per expectation.

This system can use for the complex several word based queries. The actual construction of topic space provides probable outcomes to handle the complex multiple word-based queries. It will leave for its future work.

This paper is overviews of different methods that are used for user preference forecast. They propose a personalized search model for single word queries to assist users in getting access to their interested images by predicting the searcher's preference on returned images. Now-days users of web create lots of data, and also produce huge quality of metadata. This metadata in the type of tag and social networks, groups to which they submit images. Efficiently utilize this rich user metadata in the social sharing websites for personalized search is demanding and challenging task as well as important to merit attention.

References

[1] B. Smyth, "A community-based approach to personalizing web search," *Computer*, vol. 40, no. 8, pp. 42–50, 2007.

[2] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in *Proc. SIGIR*, 2008, pp. 155–162.

[3] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'El, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized social search based on the user's social network," in *Proc. CIKM*, 2009, pp. 1227–1236.

[4] Y.Cai and Q.Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," in *Proc. CIKM*, 2010, pp. 969–978.

[5] D. Lu and Q. Li, "Personalized search on flickr based on searcher's preference prediction," in *Proc. WWW*, 2011, pp. 81–82, companion volume.

[6] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search." in *Proc. WSDM*, 2008, pp. 195–206.

[7] S. Bao, G.-R. Xue, X.Wu, Y. Yu, B. Fei, and Z. Su, "Optimizing web search using social annotations," in *Proc. WWW*, 2007, pp. 501–510.

[8] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations for information retrieval," in *Proc. WWW*, 2008, pp. 715–724.

[9] J. Tang, S. Yan, R. Hong, G. Qi, and T. Chua, "Inferring semantic concepts from community-contributed images and noisy tags," in *Proc.ACM Multimedia*, 2009, pp. 223–232.

[10] J. Tang, H. Li, G. Qi, and T. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Trans. Multimedia*, vol. 12, no. 2, pp. 131–141, Feb. 2010.

[11] M. J. Carman, M. Baillie, and F. Crestani, "Tag data and personalized information retrieval," in *Proc. SSM*, 2008, pp. 27–34.

[12] Panagiotis Symeonidis, Alexandros Nanopoulos, Yannis Manolopoulos" Tag Recommendations based on Tensor Dimensionality eduction", RecSys08, October 2325, 2008, Lausanne, Switzerland. Copyright 2008 ACM 978-1-60558- 093-7/08/10.

[13] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis, IEEE Trans. Knowl. Data Eng., vol. 22, no. 2, pp. 179192, Feb. 2010.

[14] Xiaoou Tang, Fellow, IEEE, Ke Liu, Jingyu Cui, Student Member, IEEE, Fang Wen, Member, IEEE and Xiaogang Wang, Member, IEEE," IntentSearch: Capturing User Intention for One-Click Internet Image Search " IEEE Transactions On Pattern Analysis And Machine Intelligence Vol.34 No.7 Year 2012.
[15] J.Sang, Changsheng Xu, *Senior Member, IEEE*, and Dongyuan Lu "Learn to Personalized Image Search From the Photo Sharing Websites" in IEEE Transactions On Multimedia, vol. 14, no. 4, august 2012.
[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[17] J.Jiang, D. W.Conrath Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy; The Computing Research Repository, 1997
[18] Zhongcheng Zhao, Jianzhuo Yan, Liying Fang, Pu Wang, Measuring Semantic Similarity Based On WordNet. Sixth Web Information Systems and Applications Conference. IEEE, 2009